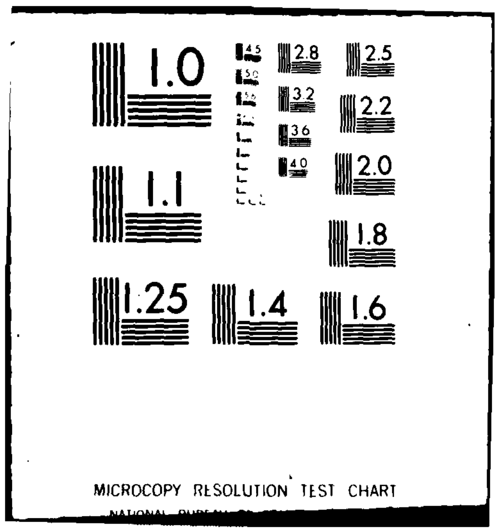
AD-A111 805 ROCHESTER UNIV NY GRADUATE SCHOOL OF MANAGEMENT F/G 12/1
ROW-CONTINUOUS FINITE MARKOV CHAINS, STRUCTURE AND ALGORITHMS.(U)
MAR 81 J KEILSON, U SUMITA, M ZACHMANN F19628-80-C-0003
UNCLASSIFIED AFGL-TR-81-0183 NL

1 of 1
AD
A111805

END
DATE
FILMED
4-82
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

(12)

AFGL-TR-81-0183

ROW-CONTINUOUS FINITE MARKOV CHAINS, STRUCTURE AND ALGORITHMS

J. Keilson
U. Sumita
M. Zachmann

University of Rochester
Graduate School of Management
Rochester, New York 14627

Scientific Report No. 3

March 1981

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

-1-

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFGL-TR-81-0183	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ROW-CONTINUOUS FINITE MARKOV CHAINS, STRUCTURE AND ALGORITHMS		5. TYPE OF REPORT & PERIOD COVERED Scientific Report No. 3
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) J. Keilson U. Sumita M. Zachmann		8. CONTRACT OR GRANT NUMBER(s) F19628-80-C-0003
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Rochester Graduate School of Management Rochester, New York 14627		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62101F 667009AG
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Monitor/LYD/Irving I. Gringorten		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 44
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Research was conducted at the M.I.T. Laboratory for Information and Decision Systems with partial support provided by the U.S.A.F. OSR Grant Number AFOSR-79-0043, U.S. Air Force Geophysics Laboratory Grant Number F19628-80-C-0003, and the National Science Foundation Grant Number NSF/ECS 79-19880.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

finite Markov chains
continuous time processes
bivariate distributions
ergodic distributions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For any finite bivariate Markov chain $[J(t), N(t)]$ on state space $\{(j,n) | 0 \le j \le J, 0 \le n \le N\}$ for which row-continuity is present, i.e., $N(t)$ changes by at most one at transitions, the ergodic distribution and mean passage times may be found by a simple algorithm. Related structure will be described. The procedure is based on probabilistic insights associated with semi-Markov processes and birth-death processes. The resulting algorithms enable efficient treatment of chains with as many as 5000 = 50 x 100 states

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

or more. Such bivariate chains are of importance to countless applied models in congestion theory, inventory theory, computer design, etc. The algorithm developed is to be used as a basis for calculating the distribution of the maximum of certain stationary meteorological processes over a specified interval.

§0. Introduction and Summary

A variety of Markov chains in continuous time modeling stochastic systems of applied interest have for their state space a rectangular lattice of states $\mathcal{B} = \{(j,n) : 0 \leq j \leq J,\ 0 \leq n \leq N\}$. When the number of states $(J+1)(N+1)$ is large, say ≥ 100, evaluation numerically of the ergodic distribution, and moments of exit times and entrance times to subsets of interest is costly and simulation is often resorted to.

For many such chains, changes in column index j or row index n at transition epochs have values $0, \pm 1$. The chains may then be described as column-continuous and row-continuous respectively. When such row-continuity is present, for example, systematic treatment of the row subsets of states as probabilistic entities provides a theoretical basis for the discussion of the chain, and algorithms for the description of the chain involving matrices of order J+1 rather than (J+1)(N+1), better suited to the capacity constraints of computers. The procedure may therefore be described as rank reducing.

Algorithms based on such treatments of rows as entities have been developed by M. Neuts [10,11], when N is infinite, for the study of queues with service times or interarrival times describable in terms of "phases". His algorithms deal with state spaces for which $N=\infty$ and the transition rates for the chain are independent of row index n, except near the boundary $n=0$. His methods are oriented largely toward the ergodic behavior of such chains.

The present treatment is primarily directed towards finite markov chains with transition rates dependent on both j and n. Entrance and exit time moments are obtained, along with the ergodic distributions.

Entry and exit time distributions, obtained via the Laguerre transform [8], will be discussed elsewhere.

It should be emphasized that the row or column orientation, natural for some systems, may be an effective tool for chain description even when no natural row or column meaning is present.

In the first section, the basic bivariate process is described and notation is developed. Several motivating examples are given. Subsequent sections develop the methodology, and algorithmic procedures, and discuss computer efficiency. In a concluding section a tandem queue with Poisson arrivals, exponential service of different rates, multiple servers, and finite waiting rooms is presented.

§1. The Bivariate Markov Process $\underline{B}(t) = [J(t),N(t)]$

Consider a bivariate Markov process $\underline{B}(t) = [J(t),N(t)]$ on $\mathcal{B} = \mathcal{J}\times\mathcal{N}$ where $\mathcal{J} = \{j: 0 \le j \le J\}$, $\mathcal{N} = \{n: 0 \le n \le N\}$. In a typical context, the process $J(t)$ is a finite Markov chain in continuous time (independent of $N(t)$), but $N(t)$ is not Markov and depends on $J(t)$. The formalism we develop is more general, however, and does not require that $J(t)$ be Markov. Suppose that $\underline{B}(t)$ is governed by the set of hazard rates $\{\nu_{(j,n),(k,m)} : j,k\varepsilon\, \mathcal{J};\ n,m\varepsilon\mathcal{N}\}$. Of interest in this paper are irreducible finite Markov chains $\underline{B}(t)$ for which $N(t)$ is skip-free in both directions so that

$$\nu_{(j,m),(k,n)} = 0 \quad \text{if} \quad |n-m| > 1. \tag{1.1}$$

It will be convenient to work with the set of states $\{(j,n)\}$ with common row index n as an entity, and to introduce the corresponding notation $\underline{\nu}_n^+$, $\underline{\nu}_n^o$, $\underline{\nu}_n^-$ to designate the transition rate matrices of order $J+1$ by

$$\text{(a)} \quad \underline{\nu}_n^o = [\nu_{(j,n),(k,n)}] \tag{1.2}$$

$$\text{(b)} \quad \underline{\nu}_n^+ = [\nu_{(j,n),k,n+1)}]$$

$$\text{(c)} \quad \underline{\nu}_n^- = [\nu_{(j,n),(k,n-1)}].$$

We will also work with the matrix

$$\underline{\nu}_n = \underline{\nu}_n^o + \underline{\nu}_n^- + \underline{\nu}_n^+ \tag{1.3}$$

and the diagonal matrix

$$\underline{\nu}_{Dn} = \text{diag}\,[\sum_k \sum_m \nu_{(j,n),(k,m)}]. \tag{1.4}$$

In the same spirit we will employ the transition probability matrices of order J+1

$$\underline{P}_{mn}(t) = [p_{(j,m)(k,n)}(t)] \;;\quad p_{(j,m),(k,n)}(t) = P[\underline{B}(t) = (k,n)\,|\,\underline{B}(0) = (j,m)] \tag{1.5}$$

and state probability row vectors

$$\underline{p}_n^T(t) = [p_{(j,n)}(t)] \;;\quad p_{(j,n)}(t) = P[\underline{B}(t) = (j,n)]. \tag{1.6}$$

The ergodic row vectors will be designated by

$$\underline{e}_n^T = [e_{(j,n)}] \tag{1.7}$$

where $e_{(j,n)} = \lim_{t\to\infty} p_{(j,n)}(t)$. Laplace transforms will often be used, with the notation typical of subsequent usage,

$$\underline{\pi}_n^T(s) = L[\underline{p}_n^T(t)] = \int_0^\infty \underline{p}_n^T(t)e^{-st}\,dt. \tag{1.8}$$

We will also employ the notion of a matrix p.d.f. [5].

<u>Def. 1.9</u> A matrix function $\underline{f}(\tau) = [f_{mn}(\tau)]$ is a matrix p.d.f. iff

a) $f_{mn}(x) \geq 0 \qquad \forall_{m,n,x}$

b) $\sum_k \int_{-\infty}^{\infty} f_{jk}(x)dx = 1 \qquad 0 \leq j \leq J.$

Such matrices play an important role in processes defined on a Markov chain [7] and in Markov renewal processes. In our setting $f_{mn}(x) = 0$ for $x < 0$.

Examples

To motivate the analysis that follows and indicate the prevalence of such row-continuous processes some examples are appropriate.

A. Contiguous Processes

Any chain $\underline{B}(t) = [J(t),N(t)]$ on $\mathcal{B} = \{(j,n) \mid 0 \leq j \leq J,\ 0 \leq n \leq N\}$ for which $\nu_{(j,n),(k,m)} = 0$ if $|j-k| > 1$ or $|m-n| > 1$ will be called contiguous. The row-continuous chains are more general in that the marginal row process need not be skip-free.

For all such contiguous processes the transition rate matrices (1.2) are tri-diagonal (cf Fig. 1.A). All such processes that are irreducible are amenable to the methods we describe.

B. Contiguous Horizontal-Vertical Processes

A subset of the contiguous processes are those for which either $J(t)$ or $N(t)$ can change, but not both simultaneously. If, for example, $J(t)$ and $N(t)$ were independent truncated birth-death processes, $\underline{B}(t)$ would be horizontal-vertical, since the probability of simultaneous change would be zero.

Another set of processes $\underline{B}(t) = [J(t),N(t)]$ has $J(t)$ an independent truncated birth-death process and $N(t)$ a dependent birth-death-like process for which the upward and downward transition rates change when $J(t)$ changes. One example of this type is a communications link carrying both voice and data [1]. We note that for such processes $\underline{\underline{\nu}}_n^+$ and $\underline{\underline{\nu}}_n^-$ are diagonal. See Fig. 1.B.

C. Tandem Queues with Blocking

A contiguous process of interest is the tandem two station series queue where each station has finite waiting room, arrivals are poisson with

rate λ, and service times are exponential with service rates η_1, η_2. For this process $\underline{\nu}_n^+$ is diagonal, $\underline{\nu}_n^-$ is upper diagonal, and $\underline{\nu}_n^o$ is lower diagonal. (A matrix $\underline{a} = [a_{ij}]$ is upper diagonal if $a_{ij} \neq 0$ implies $j = i+1$.) See Fig. 1.C.

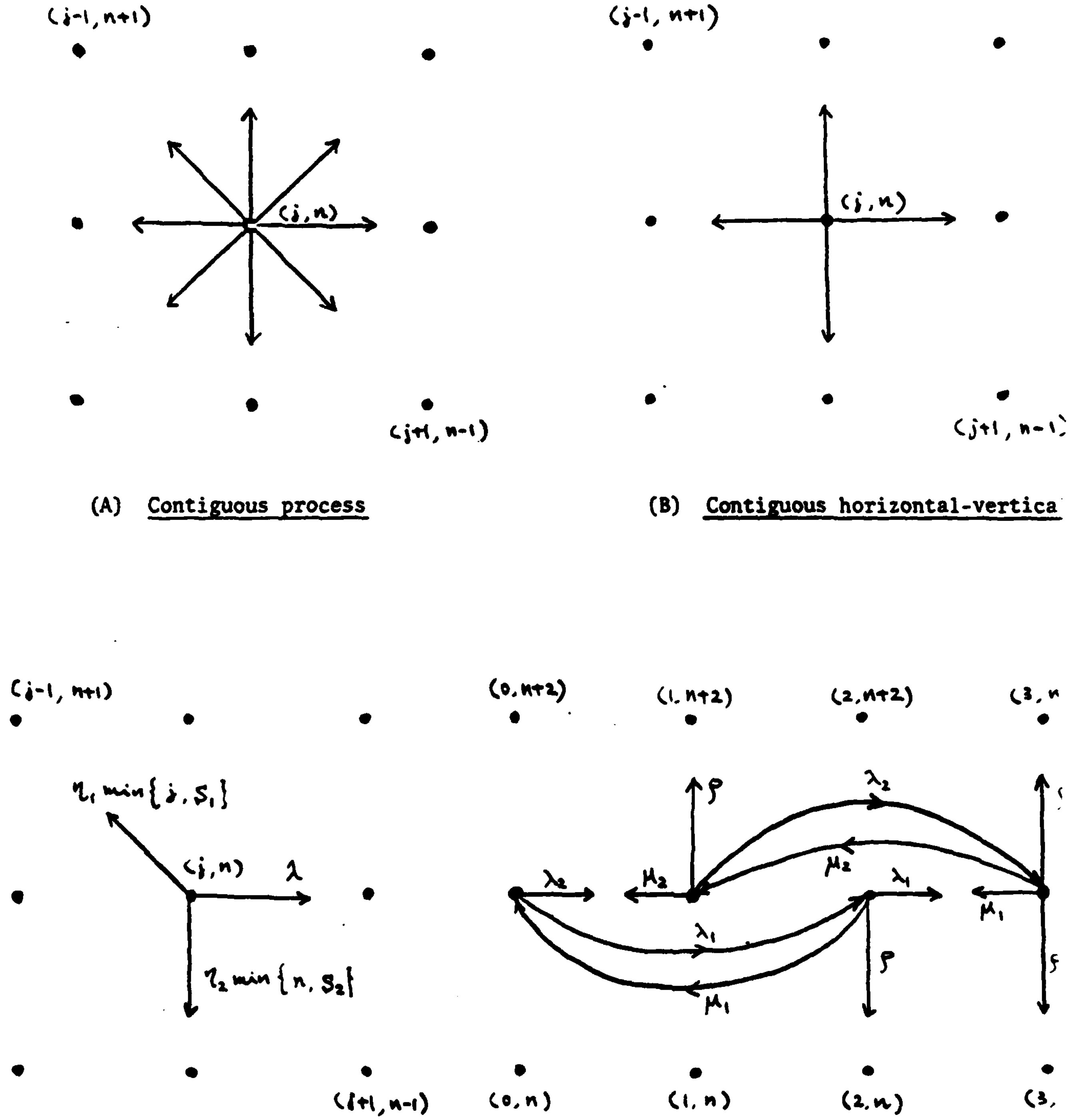
The methodology treated here allows each station to consist of a finite number of servers. Various types of blocking and feedback could as easily be analyzed, although the example of §7 has blocking defined by, the first queue stops serving while the second waiting room is full.

Extensive numerical analysis and further discussion of this tandem queue model can be found in §7.

D. Assembly Line

One interesting non-contiguous process is an assembly line with two machines in sequence and finite buffer storage between. The marginal process J(t) describes fluctuating working and non-working states of the two machines, M_1 and M_2 which are governed by failure rates μ_1, μ_2 and repair rates λ_1, λ_2 respectively. The two machines process work at the same rate of speed, i.e. have a common hazard rate ρ for completing their operation. When the second machine is down, items accumulate in the buffer. When the first machine is down, the flow of incoming items is cut off and the second machine goes idle after the buffer is emptied. The first machine stops when the buffer is full. N(t) is the number of half-finished items in the buffer.

For this model, both $\underline{\nu}_n^+$ and $\underline{\nu}_n^-$ are diagonal with two non-zero elements. The transition rate matrix $\underline{\nu}_n^o$ has a zero diagonal, and $(\underline{\nu}_n^o)^T$ has a zero diagonal. See Fig. 1.D.



(A) Contiguous process

(B) Contiguous horizontal-vertica

(C) Tandem queue

(D) Assembly line - Non-contiguous (full row shown)

Fig. 1.1 Possible transitions from an interior point

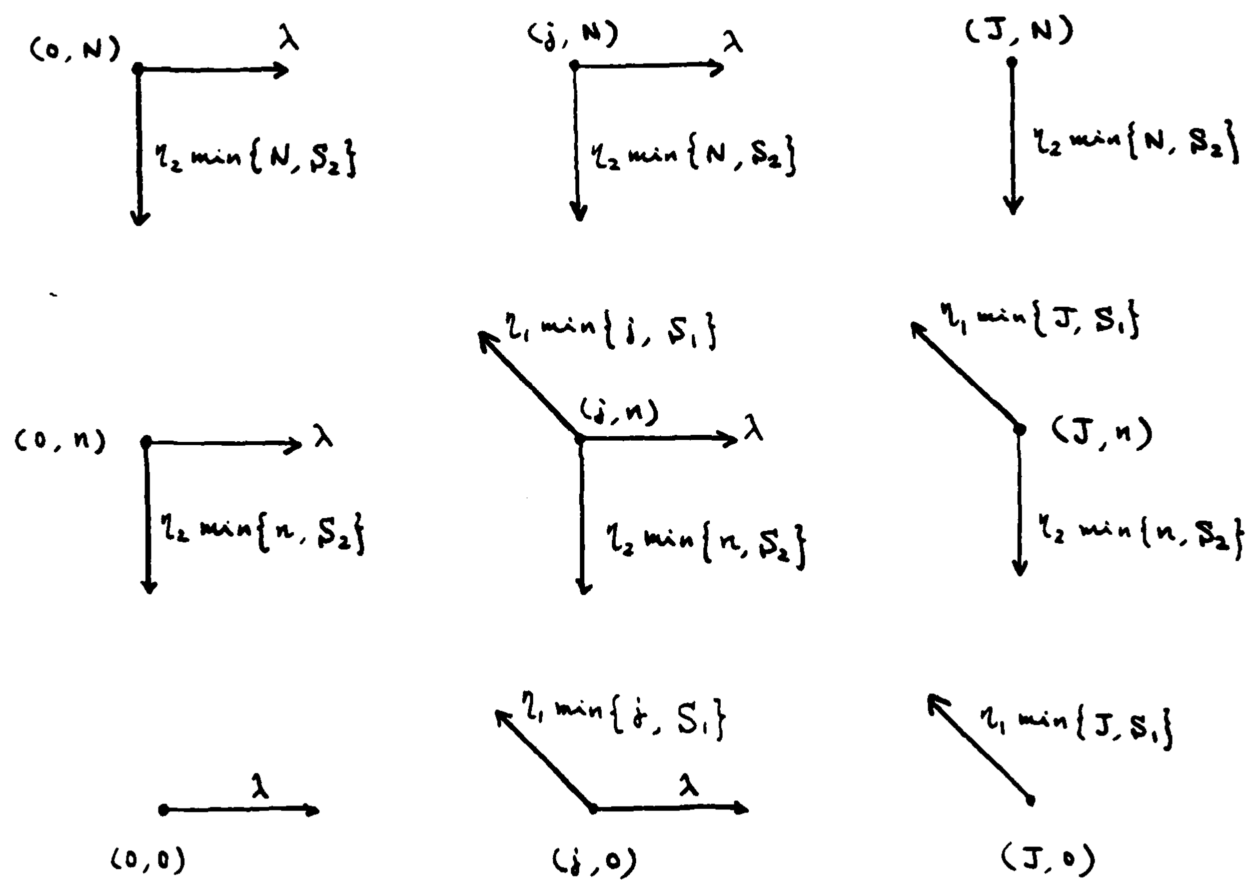


Fig. 1.1.(c) (expanded) Tandem queue

§2. Passage Time Densities

The description of the ergodic and dynamical behavior we will develop is based upon the passage time densities for the bivariate process. To exploit the skip-free feature of row-continuous processes, the passage times between states of adjacent rows are treated as building blocks. The focus on a row of states as an entity then gives rise to a matrix probability density function describing the joint distribution between the time of arrival at the adjacent row and the state reached, as a function of state of origin. From the passage time p.d.f.'s for adjacent rows one can then describe the regeneration times for the states of a row and hence the ergodic probabilities. This will be clearer from the details of the development.

To discuss the passage time densities it will be useful to work with the lossy process $J_n^*(t)$ *for row n on the set* $\{(j,n)\}$ for which other states of $\mathcal{B}$ (i.e., other rows) are absorbing. Let

$$p^*_{(j,n),(k,n)}(t) = P[J(t) = k,\ N(t') = n,\ 0 \le t' \le t \mid J(0) = j,\ N(0) = n]$$

and as for (1.5) let $\underline{p}^*_{nn}(t) = [p^*_{(j,n),(k,n)}(t)]$. Then since $\frac{d}{dt}\underline{p}^*_{nn}(t) = -\underline{p}^*_{nn}(t)\underline{\nu}_{Dn} + \underline{p}^*_{nn}(t)\underline{\nu}^o_n$ we may rewrite $\underline{p}^*_{nn}(t)$ in the form

$$\underline{p}^*_{nn}(t) = \exp\{\underline{Q}^*_n t\}\ , \tag{2.1}$$

where $\underline{Q}^*_n$ is the $\underline{Q}$ matrix for the lossy process, i.e.,

$$\underline{Q}^*_n = -\underline{\nu}_{Dn} + \underline{\nu}^o_n\ . \tag{2.2}$$

Note that $\underline{p}^*_{nn}(t)$ is strictly substochastic. From (2.1) one then has for

$\underline{\pi}^*_{nn}(s) = \mathcal{L}[\underline{p}^*_{nn}(t)]$

$$\underline{\pi}^*_{nn}(s) = \int_0^\infty e^{-st}\, \underline{p}^*_{nn}(t)dt = [s\,\underline{I} - \underline{Q}^*_n]^{-1}, \quad s > 0. \tag{2.3}$$

The chain $\underline{B}(t)$ is uniformizable [6] in the sense that there exists a ν such that $\infty > \nu > \max_{(j,m)} \sum_{(k,n)} \nu_{(j,m),(k,n)}$ when J and K are finite, that is $\mathcal{B}$ is finite. If we let

$$\underline{a}^+_n = \frac{1}{\nu}\underline{\nu}^+_n \;;\quad \underline{a}^-_n = \frac{1}{\nu}\underline{\nu}^-_n \;;\quad \underline{a}^o = \underline{I} - \frac{1}{\nu}[\underline{\nu}_{Dn} - \underline{\nu}^o_n] \tag{2.3a}$$

we have from (2.3)

$$\underline{\pi}^*_{nn}(s) = [(s+\nu)\underline{I} - \nu\underline{a}^o_n]^{-1}. \tag{2.4}$$

To proceed further, we require the passage time densities and some associated notation. Let $s^+_{n;jk}(t)$ be the joint probability density that the set $\{(j,n+1) : j\varepsilon\mathcal{J}\}$ is visited for the first time at (k,n+1) and that the time of visit is τ, so that $\sum_k \int_0^\infty s^+_{n;jk}(t)d\tau = 1$. Similarly, let $s^-_{n;jk}(\tau)$ be defined with respect to visiting the set $\{(j,n-1) : j\varepsilon\mathcal{J}\}$ for $n \geq 1$. The irreducibility of $\underline{B}(t)$ implies that (cf def. 1.9) $\underline{s}^+_n(t) = [s^+_{n;jk}(t)]$ and $\underline{s}^-_n(t) = [s^-_{n;jk}(t)]$ are matrix p.d.f.'s. Correspondingly, $\underline{s}^+_n = \int_0^\infty \underline{s}^+_n(t)dt$ and $\underline{s}^-_n = \int_0^\infty \underline{s}^-_n(t)dt$ are stochastic matrices. In terms of the Laplace transforms, $\underline{\sigma}^+_n(s) = L[\underline{s}^+_n(t)]$ and $\underline{\sigma}^-_n(s) = L[\underline{s}^-_n(t)]$, and moment matrices $\underline{\mu}^+_n = \int_0^\infty t\, \underline{s}^+_n(t)dt$, $\underline{\mu}^-_n = \int_0^\infty t\, \underline{s}^-_n(t)dt$ one has

$$\underline{s}^+_n = \underline{\sigma}^+_n(0), \quad \underline{\mu}^+_n = -\frac{d}{ds}\underline{\sigma}^+_n(s)\Big|_{s=0} \tag{2.5a}$$

$$\underline{s}^-_n = \underline{\sigma}^-_n(0), \quad \underline{\mu}^-_n = -\frac{d}{ds}\underline{\sigma}^-_n(s)\Big|_{s=0}. \tag{2.5b}$$

The skip-free feature of the row-continuous processes provides a recursion between the upward passage time p.d.f.'s $\underline{s}_n^+$, and between the downward passage time p.d.f.'s $\underline{s}_n^-$, that is a direct matrix counterpart to that present for birth-death processes where $\mathcal{J} = \{0\}$. The recursion is based on an argument similar to that employed there [6,4]. For N(t) to go from n to n+1, it must either do so directly after motion on row n, or there must be a first downward transition to row n-1, a first subsequent return to row n, and then a first subsequent arrival at row n+1. A probabilistic argument based on this then gives our basic recursive equation

$$\underline{s}_n^+(t) = \underline{p}_{nn}^*(t)\underline{\nu}_n^+ + \underline{p}_{nn}^*(t)\underline{\nu}_n^- * \underline{s}_{n-1}^+(t) * \underline{s}_n^+(t) \quad , \qquad n \geq 1 \tag{2.6}$$

where the asterisk denotes convolution in time. That is, $\underline{a}(t)*\underline{b}(t) = \int_0^t \underline{a}(t-\tau)\underline{b}(\tau)d\tau$. Consequently, from (2.6) and (2.3a),

$$\underline{\sigma}_n^+(s) = \nu\underline{\pi}_{nn}^*(s)\underline{a}_n^+ + \nu\underline{\pi}_{nn}^*(s)\underline{a}_n^-\underline{\sigma}_{n-1}^+(s)\underline{\sigma}_n^+(s), \quad n \geq 1 \, . \tag{2.7}$$

If we solve (2.7) for $\underline{\sigma}_n^+(s)$ and use (2.4) we obtain

$$[\underline{I} - (\frac{\nu}{s+\nu}) \{\underline{a}_n^o + \underline{a}_n^-\underline{\sigma}_{n-1}^+(s)\}] \underline{\sigma}_n^+(s) = (\frac{\nu}{s+\nu}) \underline{a}_n^+ \, . \tag{2.8}$$

In place of (2.8) one has for n=0, from $\underline{s}_o^+(t) = \underline{p}_{oo}^*(t)\underline{\nu}_o^+$ and (2.4)

$$[\underline{I} - (\frac{\nu}{s+\nu})\underline{a}_o^o]\underline{\sigma}_o^+(s) = (\frac{\nu}{s+\nu})\underline{a}_o^+ . \tag{2.9}$$

Equations (2.8) and (2.9) can be used to generate $\underline{\sigma}_n^+(s)$ recursively. Similar equations generate $\underline{\sigma}_n^-(s)$ recursively from $\underline{\sigma}_N^-(s)$ (cf 4.2). By letting s=0 we get:

$$[\underline{\underline{I}} - \{\underline{\underline{a}}_n^o + \underline{\underline{a}}_n^- \underline{\underline{s}}_{n-1}^+\}]\underline{\underline{s}}_n^+ = \underline{\underline{a}}_n^+ \qquad n > 0 \tag{2.10}$$

and

$$[\underline{\underline{I}} - \underline{\underline{a}}_o^o]\underline{\underline{s}}_o^+ = \underline{\underline{a}}_o^+ \tag{2.11}$$

two recursion relations for calculating $\underline{\underline{s}}_n^+$. The equivalent relations for $\underline{\underline{s}}_n^-$ appear in (4.4).

Differentiation of (2.8) and (2.9) with respect to s at s=0 leads to the following relations for the mean first passage times:

$$[\underline{\underline{I}} - \underline{\underline{a}}_n^o - \underline{\underline{a}}_n^- \underline{\underline{s}}_{n-1}^+]\underline{\underline{\mu}}_n^+ = \frac{1}{\nu}[\underline{\underline{I}} + \nu\underline{\underline{a}}_n^- \underline{\underline{\mu}}_{n-1}^+]\underline{\underline{s}}_n^+ \tag{2.12}$$

and

$$[\underline{\underline{I}} - \underline{\underline{a}}_o^o]\underline{\underline{\mu}}_o^+ = \frac{1}{\nu}\underline{\underline{s}}_o^+ . \tag{2.13}$$

We note that $\underline{\underline{s}}_n^+$ is stochastic and $\underline{\underline{a}}_n^o + \underline{\underline{a}}_n^- + \underline{\underline{a}}_n^+$ is stochastic. Hence, if $\underline{\underline{a}}_n^+$ is not zero (guaranteed by irreducibility) $\underline{\underline{a}}_n^o + \underline{\underline{a}}_n^- \underline{\underline{s}}_{n-1}^+$ is strictly substochastic, its spectral radius is less than one, and $[\underline{\underline{I}} - \underline{\underline{a}}_n^o - \underline{\underline{a}}_n^- \underline{\underline{s}}_{n-1}^+]$ is invertible. One may therefore obtain $\underline{\underline{s}}_n^+$ from (2.10) and $\underline{\underline{\mu}}_n^+$ from (2.23). We also have that $\underline{\underline{0}} \leq \underline{\underline{\sigma}}_n^+(s) \leq \underline{\underline{s}}_n^+$ for s>0 real, hence (with $\nu/s+\nu \leq 1$) we find that $[\underline{\underline{I}} - (\frac{\nu}{s+\nu})\ \{\underline{\underline{a}}_n^o + \underline{\underline{a}}_n^- \underline{\underline{\sigma}}_{n-1}^+(s)\}]$ is invertible for $s \geq 0$ real. Therefore (2.8) may be used to obtain $\underline{\underline{\sigma}}_n^+(s)$.

Of frequent interest in applications is the matrix passage time p.d.f. $\underline{\underline{s}}_{on}(\tau)$ describing the joint distribution of the time at which row n is first reached and the state reached given a start at row 0. Specifically $\underline{\underline{s}}_{on}(\tau) = [s_{on;ij}(\tau)]$ where $s_{on;ij}(\tau)$ = the joint probability density that the set $\{(j,n):j\varepsilon J\}$ is visited for the first time at (j,n) and the time of first arrival is τ, given start at (0,i). A simple probabilistic argument shows that $\underline{\underline{s}}_{on}(\tau) = \underline{\underline{s}}_o^+(\tau)*\underline{\underline{s}}_1^+(\tau)*\ldots*\underline{\underline{s}}_{n-1}^+(\tau)$. Correspondingly, $\underline{\underline{\sigma}}_{on}(s) = \underline{\underline{\sigma}}_o^+(s)\underline{\underline{\sigma}}_1^+(s)\ldots\underline{\underline{\sigma}}_{n-1}^+(s)$.

3. Ergodic Probabilities

We can now use the passage time densities of Section 2 to find the ergodic probabilities, as outlined in the introduction to that section. The basic probabilistic argument for finding the transition probabilities from the passage times goes as follows. If one is in row n at time 0, to be in row n at time t, either

a) one never left row n

b) one went to row n+1 at some first time, returned to row n for the first time subsequently, and was in row n at time t, possibly after further wanderings,

c) one went to row n-1 at some first time, etc. as in b).

Consequently, one has, for the cases $0 < n < N$, $n=0$, and $n=N$ respectively,

$$\underline{p}_{nn}(t) = \underline{p}^*_{nn}(t) + \nu\underline{p}^*_{nn}(t)\underline{a}^+_n * \underline{s}^-_{n+1}(t) * \underline{p}_{nn}(t) + \nu\underline{p}^*_{nn}(t)\underline{a}^-_n * \underline{s}^+_{n-1}(t) * \underline{p}_{nn}(t) \tag{3.1}$$

$$\underline{p}_{oo}(t) = \underline{p}^*_{oo}(t) + \nu\underline{p}^*_{oo}(t)\underline{a}^+_o * \underline{s}^-_1(t) * \underline{p}_{oo}(t) \tag{3.1a}$$

$$\underline{p}_{NN}(t) = \underline{p}^*_{NN}(t) + \nu\underline{p}^*_{NN}(t)\underline{a}^-_N * \underline{s}^+_{N-1}(t) * \underline{p}_{NN}(t) \tag{3.1b}$$

From (3.1)

$$\underline{\pi}_{nn}(s) = \underline{\pi}^*_{nn}(s) + \nu\underline{\pi}^*_{nn}(s)\underline{a}^+_n \underline{\sigma}^-_{n+1}(s)\, \underline{\pi}_{nn}(s) + \nu\underline{\pi}^*_{nn}(s)\, \underline{a}^-_n\, \underline{\sigma}^+_{n-1}(s)\, \underline{\pi}_{nn}(s) \qquad , 0 < n < N \tag{3.2}$$

so that

$$[\underline{I} - \nu\underline{\pi}^*_{nn}(s)\underline{a}^+_n\, \underline{\sigma}^-_{n+1}(s) - \nu\underline{\pi}^*_{nn}(s)\underline{a}^-_n\, \underline{\sigma}^+_{n-1}(s)]\underline{\pi}_{nn}(s) = \underline{\pi}^*_{nn}(s)\ . \tag{3.3}$$

One then finds, from (2.4), that

$$\underline{\pi}_{nn}(s) = \left(\frac{1}{s+\nu}\right)\left[\underline{I} - \left(\frac{\nu}{s+\nu}\right)\left\{\underline{a}_n^o + \underline{a}_n^+ \underline{\sigma}_{n+1}^-(s) + \underline{a}_n^- \underline{\sigma}_{n-1}^+(s)\right\}\right]^{-1}. \quad (3.4)$$

To show that the matrix expressed within brackets is in fact invertible we let

$$\underline{\beta}_n(s) = \left(\frac{\nu}{s+\nu}\right)\left\{\underline{a}_n^o + \underline{a}_n^+ \underline{\sigma}_{n+1}^-(s) + \underline{a}_n^- \underline{\sigma}_{n-1}^+(s)\right\} \quad n = 1,\ldots,N-1. \quad (3.5)$$

Note that $\underline{a}_n^o + \underline{a}_n^+ + \underline{a}_n^-$ is stochastic, and further

$$L^{-1}[\underline{\beta}_n(s)] = (\nu e^{-\nu t}\underline{I})*(\underline{a}_n^o\,\delta(t) + \underline{a}_n^+ \underline{s}_{n+1}^-(t) + \underline{a}_n^- \underline{s}_{n-1}^+(t)), \quad (3.6)$$

therefore each row of $\underline{\beta}_n(s)$ is a convex combination of rows of matrix p.d.f.'s. Hence, $\underline{\beta}_n(s)$ is itself the Laplace transform of some matrix p.d.f., say $\underline{b}_n(t)$. Hence $[\underline{I}-\underline{\beta}_n(s)]$ is invertible for s>0, real. With this identification, (3.4) becomes

$$\underline{\pi}_{nn}(s) = \left(\frac{1}{s+\nu}\right)[\underline{I}-\underline{\beta}_n(s)]^{-1} \quad n=1,\ldots,N-1. \quad (3.7)$$

In fact, with the definition

$$\underline{\beta}_o(s) = \left(\frac{\nu}{s+\nu}\right)\left\{\underline{a}_o^o + \underline{a}_o^+ \underline{\sigma}_1^-(s)\right\} \quad (3.8a)$$

and

$$\underline{\beta}_N(s) = \left(\frac{\nu}{s+\nu}\right)\left\{\underline{a}_N^o + \underline{a}_N^- \underline{\sigma}_{N-1}^+(s)\right\} \quad (3.8b)$$

we immediately get (3.7) for all $n \in N$.

To find $\underline{e}_n^T$ we must evaluate $\lim_{s\to 0^+} s\,\underline{\pi}_{nn}(s) = \underline{1}\,\underline{e}_n^T$. In [5] it is shown that

$$\lim_{s\to 0^+} s\, \underline{\pi}_{nn}(s) = \frac{1}{\nu} \cdot \frac{1}{\underline{e}_{bn}^T \underline{\mu}_{bn} \underline{1}} \cdot \underline{1}\, \underline{e}_{bn}^T \tag{3.9}$$

where $\underline{e}_{bn}^T$ is the stationary left eigenvector for $\underline{\beta}_n(0)$ and $\underline{\mu}_{bn}$ is the mean of $\underline{b}_n(t)$, that is

$$\underline{e}_{bn}^T \underline{\beta}_n(0) = \underline{e}_{bn}^T \tag{3.10}$$

and

$$\underline{\mu}_{bn} = \int_0^\infty t\, \underline{b}_n(t)dt. \tag{3.11}$$

We can easily find $\underline{\beta}_n(0)$ by setting s=0 in (3.5),(3.8a) and (3.8b), while $\underline{\mu}_{bn}$ is accessible by differentiation via

$$\underline{\mu}_{bn} = -\underline{\beta}_n'(0) = \frac{1}{\nu}\underline{\beta}_n(0) + \underline{a}_n^+ \underline{\mu}_{n+1}^- + \underline{a}_n^- \underline{\mu}_{n-1}^+ \quad n=1,\dots,N-1 \tag{3.12}$$

and

$$\underline{\mu}_{bo} = \frac{1}{\nu}\underline{\beta}_o(0) + \underline{a}_o^+ \underline{\mu}_1^- \quad ; \quad \underline{\mu}_{bn} = \frac{1}{\nu}\underline{\beta}_N(0) + \underline{a}_N^- \underline{\mu}_{N-1}^+ \, . \tag{3.13}$$

As we will see in Section 5, row balance

$$\underline{e}_{n+1}^T \underline{a}_{n+1}^- \underline{1} = \underline{e}^T \underline{a}_n^+ \underline{1} \; , \quad 0 \le n \le N-1 \tag{3.14}$$

is always present. This enables one to evaluate

$$m_{bn} \equiv \nu\, \underline{e}_{bn}^T \underline{\mu}_{bn} \underline{1} \tag{3.15}$$

recursively without computing $\underline{\mu}_{bn}$, and thus $\underline{\mu}_n^+$ and $\underline{\mu}_n^-$, in the following manner. From (3.9) one has $\underline{e}_n^T = \frac{1}{m_{bn}} \underline{e}^T_{bn}$ so that

$$\underline{e}_{n+1}^T \underline{a}_{n+1}^- \underline{1} = \frac{1}{m_{b,n+1}} \underline{e}_{b,n+1}^T \underline{a}_{n+1}^- \underline{1} \; ; \; \underline{e}_n^T \underline{a}_n^+ \underline{1} = \frac{1}{m_{b,n}} \underline{e}_{bn}^T \underline{a}_n^+ \underline{1} \, . \tag{3.16}$$

Hence, from (3.15) one sees that

$$m_{b,n+1} = m_{b,n} \frac{\underline{e}^T_{b,n+1} \underline{a}^-_{n+1} \underline{1}}{\underline{e}^T_{b,n} \underline{a}^+_{n} \underline{1}} \qquad 0 \leq n \leq N-1 \quad . \tag{3.17}$$

One may start with an arbitrary positive M_{bo} and then normalize at the end. More specifically,

$$\underline{e}^T_n = \frac{K}{m_{bn}} \underline{e}^T_{bn} , \quad m_{bo} > 0, \quad \text{arbitrary} \tag{3.18}$$

and

$$K = [\sum_{n \in N} \frac{1}{m_{bn}} \underline{e}^T_{bn} \underline{1}]^{-1} \quad . \tag{3.19}$$

4. Summary of Computation Procedure

A tabulation of the key results obtained above is given to provide an overview of the formalism, and ease of access to the formulae needed for implementation.

$$\underline{\sigma}_o^+(s) = [\underline{I} - (\frac{\nu}{s+\nu})\, \underline{a}_o^o]^{-1} (\frac{\nu}{s+\nu})\, \underline{a}_o^+ \tag{4.1}$$

$$\underline{\sigma}_n^+(s) = [\underline{I} - (\frac{\nu}{s+\nu})\, \{\underline{a}_n^o + \underline{a}_n^- \underline{\sigma}_{n-1}^+(s)\}]^{-1} (\frac{\nu}{s+\nu})\, \underline{a}_n^+$$

$$\underline{\sigma}_N^-(s) = [\underline{I} - (\frac{\nu}{s+\nu})\, \underline{a}_N^o]^{-1} (\frac{\nu}{s+\nu})\, \underline{a}_N^- \tag{4.2}$$

$$\underline{\sigma}_n^-(s) = [\underline{I} - (\frac{\nu}{s+\nu})\, \{\underline{a}_n^o + \underline{a}_n^+ \underline{\sigma}_{n+1}^-(s)\}]^{-1} (\frac{\nu}{s+\nu})\, \underline{a}_n^-$$

$$\underline{s}_o^+ = [\underline{I} - \underline{a}_o^o]^{-1}\, \underline{a}_o^+ \tag{4.3}$$

$$\underline{s}_n^+ = [\underline{I} - \{\underline{a}_n^o + \underline{a}_n^- \underline{s}_{n-1}^+\}]^{-1}\, \underline{a}_n^+$$

$$\underline{s}_N^- = [\underline{I} - \underline{a}_N^o]^{-1}\, \underline{a}_N^- \tag{4.4}$$

$$\underline{s}_n^- = [\underline{I} - \{\underline{a}_n^o + \underline{a}_n^+ \underline{s}_{n+1}^-\}]^{-1}\, \underline{a}_n^-$$

$$\underline{\mu}_o^+ = \frac{1}{\nu}\, [\underline{I} - \underline{a}_o^o]^{-1}\, \underline{s}_o^+ \tag{4.5}$$

$$\underline{\mu}_n^+ = \frac{1}{\nu}\, [\underline{I} - \{\underline{a}_n^o + \underline{a}_n^- \underline{s}_{n-1}^+\}]^{-1}\, [\underline{I} + \nu \underline{a}_n^- \underline{\mu}_{n-1}^+]\, \underline{s}_n^+$$

$$\underline{\mu}_N^- = \frac{1}{\nu}\, [\underline{I} - \underline{a}_N^o]^{-1}\, \underline{s}_N^- \tag{4.6}$$

$$\underline{\mu}_n^- = \frac{1}{\nu}\, [\underline{I} - \{\underline{a}_n^o + \underline{a}_n^+ \underline{s}_{n+1}^-\}]^{-1}\, [\underline{I} + \nu \underline{a}_n^+ \underline{\mu}_{n+1}^-]\, \underline{s}_n^-$$

$$\underline{\underline{\beta}}_n(s) = \left(\frac{\nu}{s+\nu}\right) [\underline{\underline{a}}_n^o + \underline{\underline{a}}_n^+ \underline{\underline{\sigma}}_{n+1}^-(s) + \underline{\underline{a}}_n^- \underline{\underline{\sigma}}_{n-1}^+(s)] \tag{4.7}$$

$$\underline{\underline{\beta}}_n(0) = \underline{\underline{a}}_n^o + \underline{\underline{a}}_n^+ \underline{\underline{s}}_{n+1}^- + \underline{\underline{a}}_n^- \underline{\underline{s}}_{n-1}^+$$

$$\pi_{nn}(s) = \left(\frac{1}{s+\nu}\right) [\underline{\underline{I}} - \underline{\underline{\beta}}_n(s)]^{-1} \tag{4.8}$$

$$\underline{e}_{bn}^T \underline{\underline{\beta}}(0) = \underline{e}_{bn}^T \tag{4.9}$$

$$m_{bn} = m_{b\,n-1}(\underline{e}_{bn}^T \underline{\underline{a}}_n^- \underline{1}) / (\underline{e}_{bn-1}^T \underline{\underline{a}}_{n-1}^+ \underline{1}) \quad , \quad m_{bo} > 0 \text{ arbitrary} \tag{4.10}$$

$$\underline{e}_n^T = \frac{K}{m_{bn}} \underline{e}_{bn}^T \quad ; \quad K = [\sum_{n \in N} \frac{1}{m_{bn}} \underline{e}_{bn}^T \underline{1}]^{-1} \tag{4.11}$$

§5. Row Balance and Row Generation

The computational procedure outlined in sections three and four, although a rank-reducing procedure, can be improved by using set balance on the state space, when mean passage times are not needed. A still greater improvement can be realized when the $\underline{a}_n^+$ (or $\underline{a}_n^-$) are invertible for all n. We have seen (§1) that the required invertibility is often present.

To exhibit row balance (cf. 5.9) some preliminary tools are needed.

Lemma 5.1

$$\underline{e}_{n+1}^T \underline{a}_{n+1}^- = \underline{e}_n^T[\underline{I} - \underline{a}_n^o] - \underline{e}_{n-1}^T \underline{a}_{n-1}^+ , \quad 1 \le n \le N-1 \tag{5.2}$$

and

$$\underline{e}_1^T \underline{a}_1^- = \underline{e}_o^T [\underline{I} - \underline{a}_o^o]. \tag{5.2a}$$

Proof:

The forward equations are:

$$\frac{d}{dt} \underline{p}_n^T(t) = -\nu\underline{p}_n^T(t) + \nu\underline{p}_n^T(t) \underline{a}_n^o + \nu\underline{p}_{n-1}^T \underline{a}_{n-1}^+ + \nu\underline{p}_{n+1}^T(t) \underline{a}_{n+1}^- \tag{5.3}$$

$$1 \le n \le N-1$$

and

$$\frac{d}{dt} \underline{p}_o^T(t) = -\nu\underline{p}_o^T(t) + \nu\underline{p}_o^T(t) \underline{a}_o^o + \nu\underline{p}_1^T(t) \underline{a}_1^- \tag{5.3a}$$

When we let $t \to \infty$, so that $\underline{p}_n^T(t) \to \underline{e}_n^T$, $\frac{d}{dt} \underline{p}_n^T(t) \to \underline{0}^T$ the result is immediate. □

Theorem 5.4

If $\underline{a}_n^-$ is invertible for all $1 \le n \le N$ then

$$\underline{e}_{n+1}^T = \{\underline{e}_n^T [\underline{I} - \underline{\underline{a}}_n^o] - \underline{e}_{n-1}^T \underline{\underline{a}}_{n+1}^+ \}(\underline{\underline{a}}_{n+1}^-)^{-1} \qquad 1 \leq n \leq N-1 \tag{5.5}$$

and

$$\underline{e}_1^T = \underline{e}_o^T [\underline{I} - \underline{\underline{a}}_o^o] (\underline{\underline{a}}_1^-)^{-1} \quad . \tag{5.5a}$$

Proof: trivial by Lemma 5.1. □

We see that when the ergodic distribution is desired, and this invertibility present, the $\underline{e}_n^T$ are available recursively via (5.4). This corollary implies that one need calculate only one eigenvector ($\underline{e}_{bo}^T$) in order to obtain the other $\underline{e}_n^T$ recursively.

A similar result holds when the $\underline{\underline{a}}_n^+$ for $0 \leq n < N$ are invertible. The recursion then begins with $\underline{e}_N^T$.

Theorem 5.6

If $\underline{\underline{a}}_n$ is invertible for all n>0 then

$$\underline{e}_{n-1}^T = \{\underline{e}_n^T [\underline{I} - \underline{\underline{a}}_n^o] - \underline{e}_{n+1}^T \underline{\underline{a}}_{n+1}^-\} (\underline{\underline{a}}_{n-1}^+)^{-1} \text{ , } n = 1, \ldots, N-1 \tag{5.7}$$

and

$$\underline{e}_{N-1}^T = \underline{e}_N^T [\underline{I} - \underline{\underline{a}}_N^o] (\underline{\underline{a}}_{N-1}^+)^{-1} \quad . \tag{5.7a}$$

Proof:

(5.7) is just (5.2) rearranged. The forward equation (5.3a) has its counterpart on the top row

$$\frac{d}{dt} \underline{p}_N^T(t) = - \nu\underline{p}_N^T(t) + \nu\underline{p}_N^T \underline{\underline{a}}_N^o + \nu\underline{p}_{N-1}^T(t) \underline{\underline{a}}_{N-1}^+ \quad . \tag{5.8}$$

When $t \to \infty$ (5.8) becomes (5.7a) □

Even when neither set is invertible, set balance can be used to normalize the $\underline{e}_{bn}^{T}$, as described at the end of section 3. We are now ready to prove our basic result.

Theorem 5.9 (Row balance)

$$\underline{e}_{n+1}^{T}\ \underline{a}_{n+1}^{-}\ \underline{1} = \underline{e}_{n}^{T}\ \underline{a}_{n}^{+}\ \underline{1} \tag{5.10}$$

$$\text{for } 0 \leq n \leq N-1$$

Proof:

Recall that $\underline{a}_{n}^{o}\ \underline{1} = \underline{1} - (\underline{a}_{n}^{+} + \underline{a}_{n}^{-})\underline{1}$ and $\underline{a}_{o}^{o}\ \underline{1} = \underline{1} - \underline{a}_{o}^{+}\ \underline{1}$. Combining this with Lemma 5.1 we get

$$\underline{e}_{n}^{T}\ \underline{a}_{n}^{+}\ \underline{1} - \underline{e}_{n}^{T}\ \underline{a}_{n+1}^{-}\ \underline{1} = \underline{e}_{n-1}^{T}\ \underline{a}_{n-1}^{+}\ \underline{1} - \underline{e}_{n}^{T}\ \underline{a}_{n}^{-}\ \underline{1} \tag{5.11}$$

and

$$\underline{e}_{o}^{T}\ \underline{a}_{o}^{+}\ \underline{1} = \underline{e}_{1}^{T}\ \underline{a}_{1}^{-}\ \underline{1} \tag{5.11a}$$

The proposition now follows inductively. $\square$

§ 6. The Matrix Polynomial Representation of $\underline{\sigma}_{on}(s)$

Important information on the dynamical behavior of a Markov chain is contained in the spectral structure of its first passage time densities. Knowledge of this structure, and that of the related relaxation time, is essential when one wishes to use the ergodic distribution [6].

Towards this end we introduce a representation of $\underline{\sigma}_{mn}(s)$ of the form $\underline{Q}_m(s)\underline{Q}_n^{-1}(s)$, where $\underline{Q}_r(s)$ is a matrix polynomial.

The set of poles of $\underline{\sigma}_{mn}(s)$ correspond to spectral lines.

Our approach closely parallels the work done on one-dimensional birth-death processes in [2,3] and [6]. This representation will be used to discuss the structure of $\underline{s}_{on}(\tau)$ and simple related results. We also indicate how this representation may be used, in principle, to obtain the relaxation times.

Thoerem 6.1

If $\underline{a}_n^+$ is non-singular for all n, define

$$\underline{Q}_{n+1}(s) = (\underline{a}_n^+)^{-1} \left[\left(\left(\frac{s+\nu}{\nu}\right) \underline{I} - \underline{a}_n^o\right)\underline{Q}_n(s) - \underline{a}_n^- \underline{Q}_{n-1}(s)\right] \tag{6.2}$$

with

$$\underline{Q}_o(s) = \underline{I} \quad ; \quad \underline{Q}_1(s) = (\underline{a}_o^+)^{-1} \left[\left(\frac{s+\nu}{\nu}\right) \underline{I} - \underline{a}_o^o\right]. \tag{6.2a}$$

Then

the matrix polynomial $\underline{Q}_n(s)$ is invertible $\forall\, s \geq 0$ (6.3)

$$\underline{\sigma}_n^+(s) = \underline{Q}_n(s)\, \underline{Q}_{n+1}^{-1}(s) \tag{6.4}$$

Proof: (By induction)

Clearly (6.3) is true for n=0. If we rewrite $\underline{Q}_1(s)$ as $\underline{Q}_1(s) = (\frac{s+\nu}{\nu})\ (\underline{a}_o^+)^{-1}\ [\underline{I} - (\frac{\nu}{\nu+s})\ \underline{a}_o^o]$; $\underline{a}_o^o$ is strictly substochastic hence $\underline{Q}_1(s)$ is invertible for all $s \geq 0$. Thus (6.4) makes sense and is clearly true for n=0. Now assume $\underline{Q}_n(s)$ is non-singular for $0 \leq n \leq M$ and $\underline{\sigma}_n^+(s) = \underline{Q}_n(s)\underline{Q}_{n+1}^{-1}(s)$ for $0 \leq n \leq M-1$. Then

$$\underline{Q}_{M+1}(s) = (\underline{a}_M^+)^{-1}\ [((\frac{s+\nu}{\nu})\ \underline{I} - \underline{a}_M^o)\underline{Q}_M(s) - \underline{a}_M^-\ \underline{Q}_{M-1}(s)] \tag{6.5}$$

$$= (\frac{s+\nu}{\nu})(\underline{a}_M^+)^{-1}\ [\underline{I} - (\frac{\nu}{s+\nu})\{\underline{a}_M^{\ o} - \underline{a}_M^-\ \underline{\sigma}_{M-1}^+(s)\}]\underline{Q}_M(s).$$

By the same argument as that for (2.8) et al., we get that $\underline{Q}_{M+1}(s)$ is invertible $\forall\ s \geq 0$. Postmultiplying (6.2) by $\underline{Q}_m^{-1}(s)$ and inverting one has $\underline{\sigma}_M^+(s) = \underline{Q}_M(s)\underline{Q}_{M+1}^{-1}(s)$, from (2.8). □

We note immediately that the recursion relation (6.2) implies that $(\underline{Q}_n(s))_{ij}$ is a polynomial in s of degree n. The decomposition (6.4) will be seen to be useful in a variety of ways.

The $\underline{Q}_n(s)$ arrays allow one to evaluate first passage times upwards and downwards over a number of rows rather easily. The following theorems illustrates this:

Theorem 6.6

In the context of Theorem 6.1,

a) $\underline{\sigma}_{on}(s) = \underline{Q}_n^{-1}(s)$ for $n \geq 1$

b) $Q'_{n+1}(0) = (\underline{a}_n^+)^{-1}\{\frac{1}{\nu}\ \underline{Q}_n(0) + (\underline{I} - \underline{a}_n^o)Q'_n(0) - \underline{a}_n^-\ \underline{Q}'_{n-1}(0)\}$

with $Q'_o(0) = \underline{0}$ and $\underline{Q}'_1(0) = \frac{1}{\nu}\ (\underline{a}_o^+)^{-1}$

c) $\underline{\mu}_{on} = \underline{Q}_n^{-1}(0)\ \underline{Q}_n'(0)\underline{Q}_n^{-1}(0)$

Proof:

a) $\underline{\sigma}_{on}(s) = \prod_{m=0}^{n-1} \underline{\sigma}_m^+(s) = \underline{Q}_o(s)\ (\prod_{m=1}^{n-1} \underline{Q}_m^{-1}(s)\underline{Q}_m(s))\underline{Q}_n^{-1}(s)$

$= \underline{Q}_n^{-1}(s)$ by Theorem 6.1

b) From (6.2) we get

$$\underline{Q}_{n+1}'(s) = (\underline{a}_n^+)^{-1}\ \{\frac{1}{\nu}\ \underline{Q}_n(s) + ((\frac{s+\nu}{\nu})\underline{I} - \underline{a}_n^o)\underline{Q}_n'(s) - \underline{a}_n^-\ \underline{Q}_{n-1}'(s)\} \tag{6.7}$$

At s=0 this becomes (b). The cases n=0,1 are trivial.

c) From (a) $\underline{Q}_n(s)\underline{\sigma}_{on}(s) = \underline{I}$, hence

$$\underline{Q}_n'(s)\underline{\sigma}_{on}(s) + \underline{Q}_n(s)\underline{\sigma}_{on}'(s) = 0. \tag{6.8}$$

Therefore,

$$\underline{\mu}_{on} = -\underline{\sigma}_{on}'(0) = \underline{Q}_n^{-1}(0)\underline{Q}_n'(0)\underline{Q}_n^{-1}(0)\ \square \tag{6.9}$$

Theorem 6.6 provides a computationally easy way to calculate the mean upwards passage times when the $\underline{a}_n^+$ are invertible. A dual argument eases the computation of downwards passage times when the $\underline{a}_n^-$ are invertible. The explicit formulation for arbitrary upwards passage times follows.

Theorem 6.10

a) $\underline{\sigma}_{mn}(s) = \underline{Q}_m(s)\underline{Q}_n^{-1}(s)$ for $0 \leq m \leq n \leq N$

b) $\underline{\mu}_{mn} = \underline{Q}_m(0)\underline{Q}_n^{-1}(0)\underline{Q}_n'(0)\underline{Q}_n^{-1}(0) - \underline{Q}_m'(0)\underline{Q}_n^{-1}(0)$ for $0 \leq m \leq n \leq N$

Proof:

a) Since $\underline{\sigma}_{om}(s)\underline{\sigma}_{mn}(s) = \underline{\sigma}_{on}(s)$ the result follows from (6.6a).

b) We use (a) to obtain $\underline{\sigma}_{mn}(s)\underline{Q}_n(s) = \underline{Q}_m(s)$, differentiating at s=0, one finds that

$$\underline{\sigma}_{mn}'(0)\underline{Q}_n(0) + \underline{\sigma}_{mn}(0)\underline{Q}_n'(0) = \underline{Q}_m'(0). \tag{6.11}$$

From (a) and the identity $\underline{\mu}_{mn} = -\underline{\sigma}_{mn}'(0)$ statement (b) follows. □
We note that $\underline{\sigma}_n^+(s) = \underline{Q}_n(s)\underline{Q}_{n+1}^{-1}(s)$ implies that

$$\underline{s}_n^+ = \underline{Q}_n(0)\underline{Q}_{n+1}^{-1}(0). \tag{6.12}$$

Thus, calculation of $\underline{Q}_n(0)$, and $\underline{Q}_n'(0)$ is sufficient to give $\underline{s}_{mn}$ and $\underline{\mu}_{mn}$ directly (i.e. non-recursively, and efficiently).

The matrices $\underline{Q}_n(0)$ and $\underline{Q}_n'(0)$ can be calculated recursively using (6.2) at s=0, and (6.6b). Knowledge of the matrices allows us to evaluate the mean ergodic exit time and mean stationary sojourn time, two useful measures of the dynamical behavior of the queue [6]. Both are defined with respect to a partition of the state space into two disjoint connected sets, called the good set and the bad set.

The ergodic exit time is the time required to leave the good set

given that the system has settled down, i.e. is at ergodicity (and, of course, that it is in the good set). Thus, the mean ergodic exit time, when

$$G_m = \{(j,n) : n < m\} \quad , \quad B_m = \{(j,n) : n \geq m\} \tag{6.13}$$

is given by

$$\mu_{Em} = \text{Ergodic exit time from } G_m \tag{6.14}$$

$$= \sum_{n<m} \underline{e}_n^T \underline{\mu}_{nm} \underline{1} \Big/ \sum_{n<m} \underline{e}_n^T \underline{1}$$

since entry into B_m is at row m. Using Theorem 6.10, we get

$$\mu_{Em} = \frac{\sum\limits_{n<m} \underline{e}_n^T \{\underline{Q}_n(0)\underline{Q}_n^{-1}(0)\underline{Q}_n'(0)\underline{Q}_n^{-1}(0) - \underline{Q}_m'(0)\underline{Q}_n^{-1}(0)\} \underline{1}}{\sum\limits_{n<m} \underline{e}_n^T \underline{1}} \, . \tag{6.15}$$

The stationary sojourn time is the time required to first leave the good set given that the system was stationary and a transition into the good set from the bad set just occurred. We have

$$\mu_{Vm} = \text{Stationary sojourn time on } G_m$$

$$= \{\underline{e}_m^T \underline{a}_m^- \underline{\mu}_{m-1}^+ \underline{1}\} \Big/ \underline{e}_m^T \underline{a}_m^- \underline{1} \, . \tag{6.16}$$

This can be rewritten in a simpler form using a well known result [6] as

$$\mu_{Vm} = P_\infty(G_m) \big/ i_{B_m G_m} \tag{6.17}$$

where $P_\infty(G_m)$ is the total ergodic probability on the set G_m, and $i_{B_m G_m}$ is the ergodic probability flow from B_m to G_m. This is

$$\underline{\mu}_{Vm} = \sum_{n<m} \underline{e}_n^T \underline{1} / \nu \, \underline{e}_m^T \underline{a}_m^- \underline{1} \tag{6.18}$$

For the ergodic exit time to B_m when $P_\infty(B_m) = \sum_m^N \underline{e}_n^T \underline{1} << 1$ as is often present for systems with adequate capacity or high reliability only the mean exit time is required. This may be obtained from a line search of Det $(\underline{Q}_m(s))$ along the negative real s-axis. Discussion of this is postponed to a subsequent paper.

§7. A Tandem Queue Example

A tandem queue is a queueing system with distinct service facilities connected in series, i.e. the customer output stream of the first facility is the input stream of the second [9]. To illustrate the algorithmic procedure we have developed, one such tandem queue will be analyzed. The example selected has been chosen primarily for didactic reasons. A more complex and realistic example could have been analyzed with little increase in machine cost.

The tandem queue evaluated is a two server series queue with blocking and finite buffers (see §1.E). The first service facility has 8 buffer slots and 4 servers. The second facility has 4 buffer slots and 5 servers. A flow diagram is given in Figure 7.1. The corresponding rate matrices are given in Figure 7.2. When the first queue is full customers balk and go elsewhere. When the second queue is full, the first queue stops serving until a space at the second facility opens up. The model can easily be modified to allow different types of blocking, and features such as feedforward, feedback, etc.

In Figure 7.1 the occupancy level at the first facility is given by the coordinate j which corresponds to the number of people in service or waiting there. The occupancy level at the second facility is given by the coordinate n. The blocking may be seen in the absence of the transition rates associated with μ_1 on the top row.

Figure 7.2 displays the three matrices $\underline{\nu}_n^o$, $\underline{\nu}_n^+$, and $\underline{\nu}_n^-$ when n=3. The matrix $\underline{\nu}_n^o$ is upper diagonal (λ) for all n. The only transitions which have no impact on the 2nd queue are the arrivals (at rate λ) to the first queue, which increase that occupation level by 1. The matrix

$\underline{\nu}_n^-$ is diagonal, with diagonal elements $\mu_2 \cdot \min\{n,k_2\}$ where k_2 is the number of servers in the 2nd service facility. Finally, $\underline{\nu}_n^+$ is lower diagonal. Increases in queue length at the 2nd service facility are caused by departures from the first facility.

The results are shown in Figs. 7.3 (a),(b),(c), 7.4 and 7.5 for a traffic intensity of 0.4 at the first facility and 0.6 at the second. Figs. 7.3 (a),(b),(c) describe the ergodic probabilities as a function of (j,n). An examination of these figures immediately reveals two features. First, the ergodic probabilities are signficant for modest values of j and n, arising from the moderate traffic intensities. The small ridge visible along the upper edge of Fig. 7.3 (a) indicates some blocking.

The ergodic probabilities may be useless when system parameters change before ergodicity is reached. The information contained in the first passage times may then be helpful. The ergodic distribution, for example, might cause concern over large probabilities of saturation, blocking, long waiting times and so on. If, however, we look (cf. Fig. 7.5) at the mean first passage times from idleness, we see that the mean time to saturation is on the order of 4 days (when our time scale is in hours). If this queue modeled a system that starts anew each day, one would be less inclined to worry.

The tables in Figure 7.4 give, for differing levels, the mean sojourn time [6], i.e., the mean time spent above a certain level after the level is reached. This gives a feeling, for example, for the persistence time in a congested state. The table 7.5 presents

mean first passage times one step upwards (downwards). The (j,n) entry in the table is the mean time to go from (j,n) to a state in the row n+1 (n-1). Again, the effect of blocking is noticeable in the dropoff (for the upward table) of the passage time with increasing values of j.

Ergodic distributions for the same system with different parameters are shown in Figure 7.6 for comparison purposes.

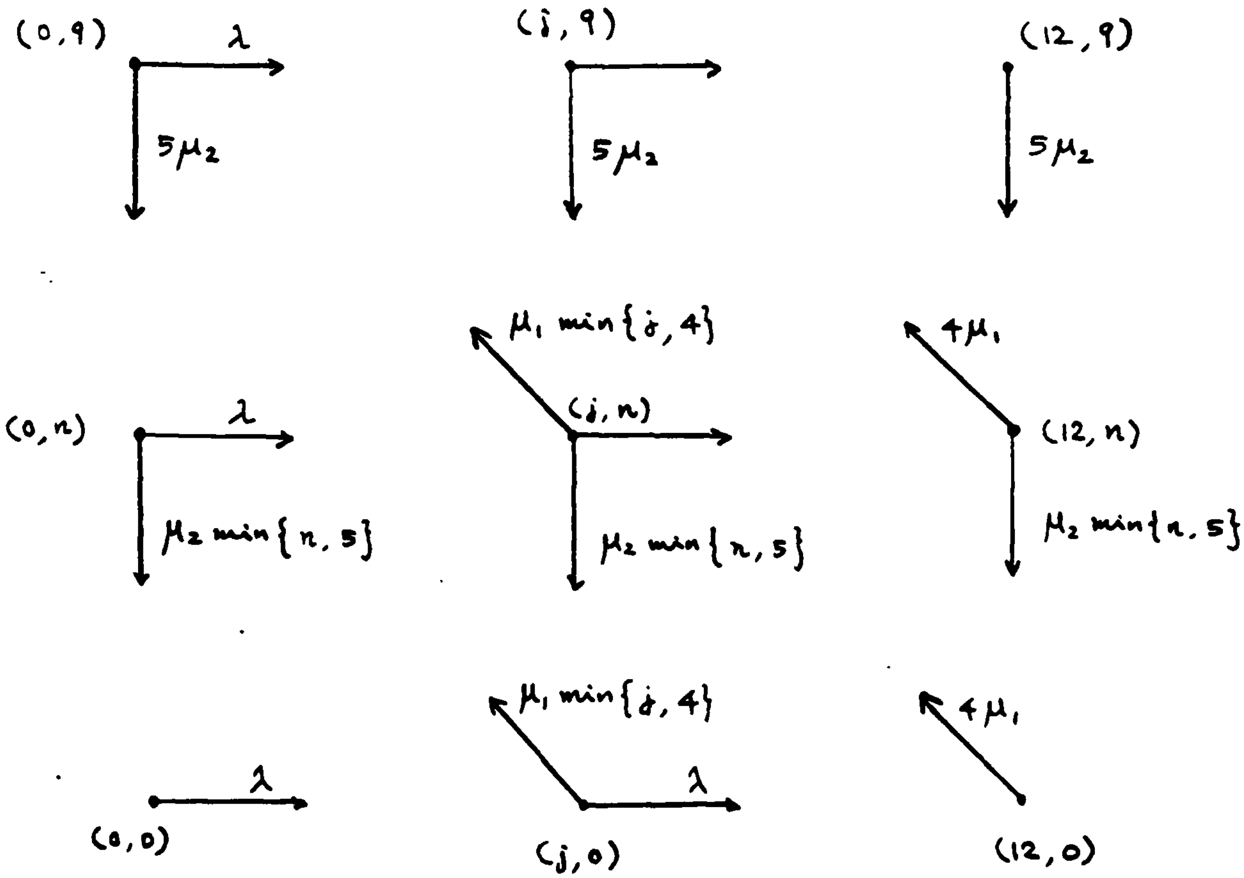


Fig. 7.1 Transition rate structure for the tandem queue

Transition rates are shown for representative edge, corner, and interior states.

$$\underline{\underline{\nu}}_n^{\,0} = \begin{bmatrix} 0 & \lambda & & & & & \\ & 0 & \lambda & & & \underline{\underline{0}} & \\ & & 0 & \ddots & & & \\ & & & \ddots & \ddots & & \\ & \underline{\underline{0}} & & & \ddots & \lambda & \\ & & & & & 0 & \lambda \\ & & & & & & 0 \end{bmatrix} \qquad 13 \times 13$$

$$\underline{\underline{\nu}}_3^{\,-} = \begin{bmatrix} 3\mu_2 & & & & \\ & 3\mu_2 & & \underline{\underline{0}} & \\ & & 3\mu_2 & & \\ & \underline{\underline{0}} & & \ddots & \\ & & & & 3\mu_2 \end{bmatrix} \qquad 13 \times 13$$

$$\underline{\underline{\nu}}_3^{\,+} = \begin{bmatrix} 0 & & & & & & \\ \mu_1 & 0 & & & \underline{\underline{0}} & & \\ & 2\mu_1 & 0 & & & & \\ & & 3\mu_1 & 0 & & & \\ & & & 4\mu_1 & 0 & & \\ & \underline{\underline{0}} & & & \ddots & \ddots & \\ & & & & & 4\mu_1 & 0 \end{bmatrix} \qquad 13 \times 13$$

Fig. 7.2 Transition rate matrices for the tandem queue model (n=3)

TANDEM QUEUE EXAMPLE

SIZES: *BUFFER* (1) = 08 *BUFFER* (2) = 04 *SERVERS* (1) = 04 *SERVERS* (2) = 05

ARRIVAL RATE = 4.000 *SERVICE RATE* (1) = 2.500 *SERVICE RATE* (2) = 1.300

ERGODIC PROBABILITIES

i ↓ J →	0	1	2	3	4	5	6	7	8	9	10	11	12
9	0.0017	0.0038	0.0047	0.0042	0.0031	0.0020	0.0013	0.0008	0.0005	0.0003	0.0002	0.0001	0.0001
8	0.0038	0.0069	0.0067	0.0046	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000
7	0.0069	0.0117	0.0101	0.0060	0.0028	0.0014	0.0008	0.0004	0.0003	0.0001	0.0001	0.0001	0.0000
6	0.0118	0.0192	0.0158	0.0088	0.0038	0.0017	0.0008	0.0004	0.0002	0.0001	0.0001	0.0000	0.0000
5	0.0194	0.0313	0.0254	0.0138	0.0056	0.0024	0.0010	0.0005	0.0002	0.0001	0.0001	0.0000	0.0000
4	0.0317	0.0509	0.0410	0.0220	0.0089	0.0036	0.0015	0.0006	0.0003	0.0001	0.0001	0.0000	0.0000
3	0.0413	0.0662	0.0531	0.0284	0.0114	0.0046	0.0019	0.0008	0.0003	0.0001	0.0001	0.0000	0.0000
2	0.0403	0.0645	0.0516	0.0276	0.0111	0.0044	0.0018	0.0007	0.0003	0.0001	0.0001	0.0000	0.0000
1	0.0262	0.0419	0.0335	0.0179	0.0072	0.0029	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000
0	0.0085	0.0136	0.0109	0.0058	0.0023	0.0009	0.0004	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000

Fig. 7.3(a) The ergodic probabilities for the tandem queue model

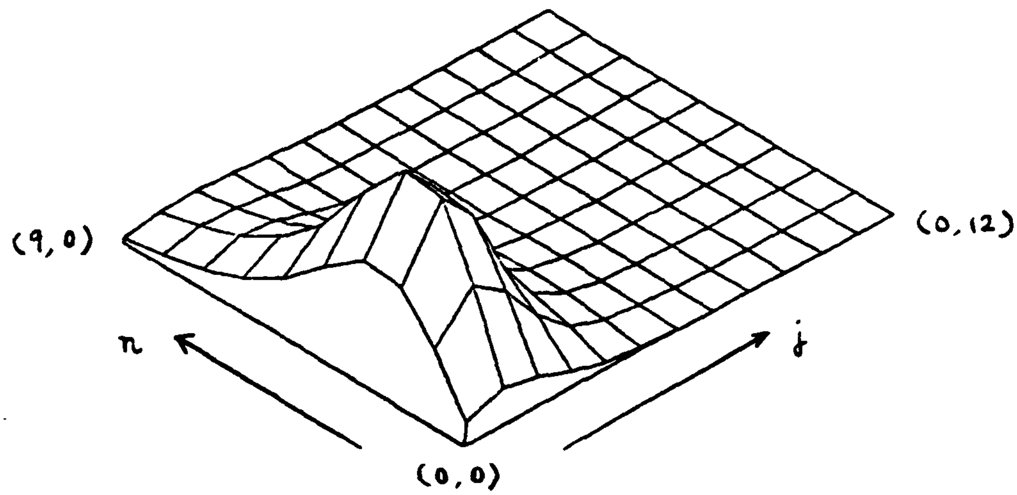


Fig. 7.3(b) The ergodic probabilities for the tandem queue model

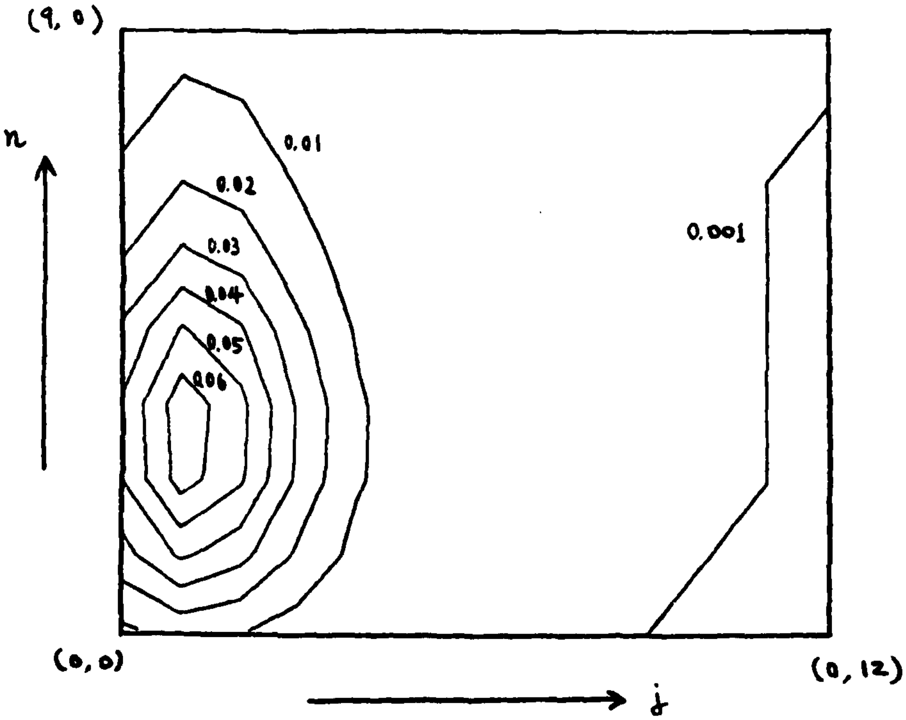


Fig. 7.3(c) The contours of the ergodic probabilities

MEAN SOJOURN TIMES ON THE SET { (J,N) : N ≥ X }

X =	1	2	3	4	5	6	7	8	9
	5.600746	1.568603	0.768045	0.496777	0.392214	0.379189	0.347591	0.277949	0.153846

MEAN SOJOURN TIMES ON THE SET { (J,N) : J ≥ X }

X =	1	2	3	4	5	6	7	8	9	10	11	12
	0.039110	0.014885	0.009001	0.007106	0.007581	0.008174	0.008832	0.009430	0.009751	0.009483	0.008210	0.005347

Fig. 7.4 The mean sojourn time for the tandem queue model

MEAN FIRST PASSAGE TIMES ONE STEP UPWARDS (↑N)

N ↓ J →	0	1	2	3	4	5	6	7	8	9	10	11	12
8	18.8602	12.3717	7.7164	4.6200	2.7606	1.9648	1.4850	1.1628	0.9357	0.7732	0.6594	0.5870	0.5534
7	11.3294	7.4297	4.6352	2.7777	1.6617	1.1814	0.8929	0.7024	0.5721	0.4824	0.4219	0.3847	0.3678
6	6.7328	4.3919	[illegible].7236	1.6223	0.9665	0.6860	0.5235	0.4225	0.3583	0.3174	0.2919	0.2771	0.2707
5	3.9476	2.5233	1.5257	0.8898	0.5278	0.3828	0.3078	0.2670	0.2444	0.2319	0.2251	0.2216	0.2202
4	2.2708	1.3542	0.7750	0.4433	0.2735	0.2186	0.1958	0.1859	0.1816	0.1797	0.1789	0.1786	0.1785
3	1.4271	0.7981	0.4451	0.2636	0.1786	0.1590	0.1531	0.1513	0.1507	0.1505	0.1505	0.1505	0.1505
2	0.9615	0.5103	0.2888	0.1846	0.1374	0.1308	0.1296	0.1294	0.1294	0.1294	0.1294	0.1294	0.1294
1	0.6831	0.3502	0.2092	0.1451	0.1150	0.1131	0.1130	0.1130	0.1130	0.1130	0.1130	0.1130	0.1130
0	0.5055	0.2555	0.1652	0.1217	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000

MEAN FIRST PASSAGE TIMES ONE STEP DOWNWARDS (↓N)

N ↓ J →	0	1	2	3	4	5	6	7	8	9	10	11	12
9	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538	0.1538
8	0.1848	0.2351	0.2839	0.3295	0.3672	0.3816	0.3871	0.3892	0.3900	0.3903	0.3905	0.3905	0.3905
7	0.1975	0.2685	0.3576	0.4581	0.5582	0.6226	0.6659	0.6950	0.7145	0.7275	0.7362	0.7417	0.7448
6	0.2033	0.2836	0.3920	0.5267	0.6748	0.7894	0.8820	0.9579	1.0202	1.0711	1.1121	1.1439	1.1648
5	0.2057	0.2900	0.4068	0.5573	0.7307	0.8766	1.0043	1.1180	1.2198	1.3107	1.3907	1.4583	1.5065
4	0.2728	0.3773	0.5185	0.6971	0.9027	1.0854	1.2535	1.4103	1.5574	1.6951	1.8224	1.9348	2.0175
3	0.4333	0.6057	0.8166	1.0622	1.3279	1.5609	1.7811	1.9925	2.1963	2.3924	2.5782	2.7457	2.8699
2	0.9427	1.3055	1.6983	2.1009	2.4898	2.8046	3.0961	3.3773	3.6516	3.9186	4.1742	4.4057	4.5761
1	4.0319	5.0923	6.0317	6.8311	7.4829	7.9339	8.3233	8.6893	9.0436	9.3862	9.7177	10.0143	10.2292

MEAN FIRST PASSAGE TIME FROM IDLE TO GIVEN N

N =	1	2	3	4	5	6	7	8	9
	0.5055	0.9382	1.4672	2.2167	3.4175	5.5906	9.3470	15.6818	26.2116

Fig. 7.5 The mean first passage times for the tandem queue model

TANDEM QUEUE EXAMPLE

SIZES: BUFFER (1) = 08   BUFFER (2) = 04   SERVERS (1) = 04   SERVERS (2) = 05

ARRIVAL RATE = 8.000   SERVICE RATE (1) = 4.000   SERVICE RATE (2) = 3.000

ERGODIC PROBABILITIES

N ↓ J→	0	1	2	3	4	5	6	7	8	9	10	11	12
9	0.0005	0.0012	0.0017	0.0016	0.0011	0.0007	0.0005	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000
8	0.0013	0.0027	0.0031	0.0024	0.0014	0.0008	0.0005	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000
7	0.0026	0.0054	0.0057	0.0040	0.0022	0.0012	0.0007	0.0004	0.0002	0.0001	0.0001	0.0000	0.0000
6	0.0051	0.0104	0.0106	0.0072	0.0037	0.0019	0.0010	0.0005	0.0003	0.0002	0.0001	0.0000	0.0000
5	0.0097	0.0196	0.0198	0.0133	0.0067	0.0034	0.0018	0.0009	0.0005	0.0002	0.0001	0.0001	0.0000
4	0.0184	0.0368	0.0370	0.0248	0.0125	0.0063	0.0032	0.0016	0.0008	0.0004	0.0002	0.0001	0.0000
3	0.0276	0.0553	0.0554	0.0370	0.0186	0.0093	0.0047	0.0024	0.0012	0.0006	0.0003	0.0002	0.0001
2	0.0311	0.0622	0.0622	0.0415	0.0208	0.0104	0.0052	0.0026	0.0013	0.0007	0.0004	0.0002	0.0001
1	0.0233	0.0466	0.0466	0.0311	0.0156	0.0078	0.0039	0.0020	0.0010	0.0005	0.0003	0.0001	0.0001
0	0.0087	0.0175	0.0175	0.0117	0.0058	0.0029	0.0015	0.0007	0.0004	0.0002	0.0001	0.0000	0.0000

SIZES: BUFFER (1) = 08   BUFFER (2) = 04   SERVERS (1) = 04   SERVERS (2) = 05

ARRIVAL RATE = 8.000   SERVICE RATE (1) = 5.000   SERVICE RATE (2) = 4.000

ERGODIC PROBABILITIES

N ↓ J→	0	1	2	3	4	5	6	7	8	9	10	11	12
9	0.0001	0.0003	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0004	0.0007	0.0006	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0011	0.0018	0.0015	0.0008	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0028	0.0046	0.0037	0.0020	0.0008	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0071	0.0114	0.0092	0.0049	0.0020	0.0008	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0178	0.0286	0.0229	0.0122	0.0049	0.0020	0.0008	0.0003	0.0001	0.0001	0.0000	0.0000	0.0000
3	0.0357	0.0571	0.0457	0.0244	0.0098	0.0039	0.0016	0.0006	0.0003	0.0001	0.0000	0.0000	0.0000
2	0.0535	0.0857	0.0685	0.0366	0.0146	0.0059	0.0023	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
1	0.0535	0.0857	0.0685	0.0366	0.0146	0.0059	0.0023	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
0	0.0268	0.0428	0.0343	0.0183	0.0073	0.0029	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000

Fig. 7.6 Ergodic probabilities of tandem queues with different parameters

§8. Rank Reduction in the Row-Continuous Model

The algorithms presented in the preceding sections exploit the block tridiagonal structure of the transition rate matrix associated with a bivariate row-continuous markov chain. These algorithms are considerably more efficient than the general algorithms for computation on a markov chain both in computer time needed to attain desired results and in storage needed to perform these computations. To illustrate this fact we compare the resources required to calculate ergodic probabilities for the tandem queue of section 7.

We note that the tandem queue is not time reversible, and so we could not have used detailed balance simplification. We also note that the mean passage times, themselves of some importance, are a byproduct of the row-continuous ergodic distribution calculations. Finally, the tandem queue does have the property that $\underline{a}_n^-$ is trivially invertible, a feature which reduces computation time.

In order to calculate ergodic probabilities efficiently we must store the $\underline{a}_n^-$, $\underline{a}_n^+$, $\underline{a}_n^o$, of course, as well as the $\underline{s}_n^+$ and $\underline{s}_n^-$. The storage requirements are then $O(J^2N)$. For a contiguous problem (such as this) one would pick the row direction J(t) to be that with the fewest states. A naive generalized approach would store the full transition matrix, thus using up $O(J^2N^2)$ memory locations. For example, one tandem queue problem with J=8 and N=80 was evaluated in a workspace of 140,000 bytes. The transition rate matrix for the bivariate problem would require more than 4,000,000 bytes alone.

The example of section 7 was coded in APL, a notoriously inefficient language. Table 8.1 gives some times for the tandem queue evaluation.

These were run on an Amdahl 470 v/6 timesharing with 20 users. The timings include all setup and all calculations required to find the ergodic probabilities. For comparision purposes, a single eigenvalue was found for a 64 x 64 transition matrix. This required approximately 70 CPU seconds using an optimized successive approximation. No effort was made to optimize the programs used for Table 8.1. In fact, the matrices were regenerated when needed.

A complexity argument can be used to find the number of operations required to evaluate the ergodic distribution. This would severely understate the advantage of using the previous algorithms, because a general approach would require repeated paging in and out of the system parameters for a problem of any size. Nevertheless, the run-time was determined empirically as $O(J^2N)$. We expect that when it is necessary to calculate $(\underline{a}_n^-)^{-1}$, rather than being able to explicitly determine them (as in the tandem queue where $\underline{a}_n^-$ is diagonal) the calculation time would be $O(J^3N)$, still a great improvement over the general case.

J	N	CPU time (seconds on an IBM 370-158)
2	2	1.2
4	4	1.7
4	16	5.3
6	60	22
6	80	28
8	8	4.6
8	16	8.5
8	40	20.2
8	80	40
12	16	16.4
12	24	24
16	8	17.5
16	16	31

Fig. 8.1 Approximate CPU time required to calculate ergodic probabilities

Acknowledgements

The authors wish to thank S. Graves for helpful discussion at the onset of the study in the summer of 1980. They also wish to thank D. Friedman for providing impetus to the work through his interests in voice-data communications. The computer support of STSC, Inc. is gratefully acknowledged.

References

1. Friedman, D., "Queueing Analysis of A Dynamically Shared Voice-Data Link", Ph.D. Thesis, M.I.T., Department of Electrical Engineering and Computer Science, 1981 (also, LIDS Report).

2. Karlin, S. and J.L. McGregor, "The Differential Equations of Birth-and-Death Processes and the Stieltjes Moment Problem", Trans. Amer. Math. Soc., Vol. 85, pp. 489-546, 1957.

3. Karlin, S and J.L. McGregor, "The Classification of Birth and Death Processes", Trans. Amer. Math. Soc., Vol. 86, 1957.

4. Keilson, J., "A Review of Transient Behavior in Regular Diffusion and Birth-Death Processes--Part II", Journal of Applied Probability, Vol. 2, pp. 405-425, 1965.

5. Keilson, J., "On the Matrix Renewal Function for Markov Renewal Processes", The Annals of Math. Stat., Vol. 40, pp. 1901-1907, 1969.

6. Keilson, J., Markov Chain Models--Rarity and Exponentiality, Springer-Verlag, New York, 1979.

7. Keilson, J. and D.M.G. Wishart, "A Central Limit Theorem for Processes Defined on a Finite Markov Chain", Proc. of the Camb. Philo. Soc., Vol. 60, pp. 547-567, 1964.

8. Keilson, J., W. Nunn and U. Sumita, "The Laguerre Transform", Professional Paper 284, Center for Naval Analysis, 1980.

9. Kleinrock, L., Queueing Systems, Vol. 2: Computer Applications, John Wiley & Sons, New York, 1976.

10. Neuts, M.F., "Markov Chains with Applications in Queueing Theory Which Have a Matrix-geometric Invariant Vector", Adv. Appl. Prob., Vol. 10, pp. 185-212, 1978.

11. Neuts, M.F., Matrix Geometric Solutions in Stochastic Models--An Algorithmic Approach, N. Holland, 1981 (to appear).